

INFORMATION RETRIEVAL WITH NON-NEGATIVE MATRIX FACTORIZATION

ABSTRACT OF THE INVENTION

Disclosed is a method of indexing a database of documents, comprising providing
5 a vocabulary of n terms, indexing the database in the form of a non-negative $n \times m$ index
matrix V , wherein each of its m columns represents an j^{th} document having n entries
containing a function of the number of occurrences of a i^{th} term of said vocabulary
appearing in said j^{th} document, factoring out non-negative matrix factors T and D such
that $V \approx TD$, and wherein T is an $n \times r$ term matrix, D is an $r \times m$ document matrix, and r
10 $< nm/(n+m)$. The index so generated is useful in two-pass information retrieval systems.